2

11. SAMPLE SELECTION

A. Introduction

This section describes statistical sampling[1] methods used by NECA in its annual data collection program for average schedule formula development. The sampling design identifies the sample cost and average schedule companies to be used for collecting accounting and demand data for a given year. A well-designed sample provides a desired level of precision and reliability and eliminates the need to collect data from the entire population of cost and average schedule companies. By employing statistical sampling methods, NECA and pool members save time, labor, and money without sacrificing accuracy.

This average schedule study uses a five-year sample design, first introduced in the 1998 study. This sample design provides for samples of average schedule and cost study areas to supply data to NECA over the five-year period from 1998 to 2002.

Large and small ECs are distinguished according to group designations developed by NECA for use in its annual Access Tariff Filing?[2] According to this classification scheme, group A includes all Regional Bell Operating Company study areas and study areas of other large holding companies not in the NECA pools. Group B includes larger cost study area members of the NECA pools, many of which are affiliated with other study areas through holding companies.[3] Because of their size and

1 Statistical sampling is a procedure used in analytical studies to provide an estimate, with an acceptable precision, of *the true* value of a critenon variable underlying *an* entire population, but at considerable savings in time and money.

2 **See,** *e.g.* National Exchange Carrier Association, Inc., Tariff F.C.C. No. 5, Transmittal No. 939, filed June 17,2002 at Vol. 2, pp. 2 -3 (2002 Annual Access Tariff Filing).

3 Group B companies include: ALLTEL, Anchorage Telephone Utility, Century, Pacific Telecom, Puerto Rico Telephone, and Telephone and Data Systems (TDS). Some study areas owned by holding companies in the group are included in group D because they utilize average schedules.

operating characteristics, group A and B companies are not representative of average schedule companies and therefore are not asked to supply data for average schedule formula development. Group C contains smaller cost study areas that are similar to average schedule companies, and group D consists of all average schedule study areas.

In 1998, NECA developed a five-year sampling design, similar to the 1993 five-year sampling design, to draw samples for each of the five years from 1998 to 2002. In this design, NECA ensured that additional 'small' average schedule study areas were included? 'Small' study areas were defined as those with fewer than 200 access lines per exchange. The design entailed defining stratification attributes, determination of sample size, and allocation of the sample to strata, sample selection and assignment of study areas to specific data collection years. The data used to design the sample included the NECA tariff filing information that designates a study area as Group B, C or D, Traffic Sensitive pool participation status, exchange counts, provision of line haul, provision of host/remote facilities, provision of special access services, provision of tandem access facilities and total net earned interstate access revenues.

Section B describes features of NECA's 1998 five-year sampling design that meet sample data needs and enables NECA to combine samples from two consecutive years to improve precision.

4 In a December 1997 order, the Common Carrier Bureau directed NECA to work with its staff to assure that sample data used by NECA accurately reflects all sizes of average schedule companies. ***See*** NECA Proposed Modification to the 1997 Interstate Average Schedule Formulas and Proposed Further Modifications to the 1997-1998 Interstate Average Schedule Formulas, AAD 97-109, ***Order on Reconsideration and Order,*** 13 FCC Rcd 101 16 (1997) ***(December*** 1997 ***Order).*** The Accounting Safeguards Division also expressed concern that NECA's sample data was not representative of companies of all ***sizes*** in a June 1998 order. ***See*** NECA Proposed Modifications to the 1998-99 Interstate Average Schedule Formulas, AAD 98-20, ***Order,*** 13 FCC Rcd 17351 (1998) *(June* ***1998 Order).***

Section C defines the nine attributes of a study area that were used for an initial classification of the average schedule population into 39 classes and the cost company population into 52 classes. A special size criterion was included in the average schedule company classification method, to enable inclusion of proportionately smaller average schedule study areas.

Section D describes the criteria used to collect classes of study areas into sampling strata. Classes that include only a few study areas are combined with others, and classes that contain high variations in study area revenues were split into subclasses by revenue size. This procedure resulted in 14 cost study area strata and 14 average schedule study area strata. Stratification of the population is done to assure that the sample will provide the desired precision level and meet specialized data needs.

In Section E, NECA explains the determination of sample size, drawing upon statistical formulas found in sampling textbooks. The stratified sample with optimum allocation of the sample among strata helps produce statistical results with a desired level of precision at a fraction of the resource cost of examining the entire population. NECA demonstrates that its annual sample size of approximately 100 cost and 100 average schedule study areas is sufficient to ensure that the proposed formulas provide results with the desired level of precision.

Section F describes the allocation of the five-year sample size among different strata. NECA uses the "Neyman Allocation" method to determine the optimum number of study areas to be sampled from each stratum. In some strata, the optimum sample size equals or exceeds the total stratum size. In such strata, data will be collected for every study area over the five-year period, and from some more than once. In other cases the optimum sample size is less than the total stratum size. In such strata, not all study areas will submit data during the five-year period.

Section G explains random sampling of study areas from each stratum using probabilities of selection proportional to the size of each study area. This procedure called Probability Proportional to Size sampling (PPS Sampling), assigns a greater probability of selection to larger study areas. Section H explains the sample weight calculation. These weights are applied to the sample data to provide parameter estimates for the average schedule population.

Section I describes the assignment of sample study areas from each strata to sample years. **This** technique ensures that data from the larger study areas are included in every average schedule study, and that the same study **area** will not be included in the sample **for** two consecutive years, thereby spreading the cost of responding to sample data requests among more study **areas** and increasing the effective sample size for average schedule studies.

Data that underlie the 2002 Study are from the annual samples of study areas collected in 2000 and 2001. **This** section of the filing produces the list of sample study areas, listed in Appendix A1, and their sample weights, displayed in Appendix D1 and D2, that were used in the 2002 Study.

B. Five-Year Sampling Design

The five-year sampling design selects a five-year sample, and then assigns members of the sample to data collection years.[5] A five-year sampling design methodology was developed in 1998 to support average schedule study activities for the 1999-2003 period. It is similar to the five-year sampling methodology developed in 1993 to support average schedule study activities for the 1994-1998 period.[6]

---

5 NECA introduced the first multi-year design method in 1988, which supported average schedule studies between 1989 and 1993. ***See, e.g.*** National Exchange Carrier Association, Inc., 1990 Modification of Average Schedules, December 29, 1989.

6 **See,** *e.g.* National Exchange Carrier Association, Inc., 1995 Modification of Average Schedules, December 30, 1994.

NECA's five-year sampling design plans for samples of cost and average schedule study areas to supply data to **NECA** in each year within a five-year period. **NECA** finds **this** plan an effective method because it achieves a targeted precision level while fairly distributing reporting burdens among companies. The plan uses an annual sample **size,** which is sufficient to maintain the desired precision level as the population changes over the five-year period. To protect against possible degradation in precision level, **NECA** redesigns the sample to reflect the current population every five years.

Use of a five-year sampling design allows **NECA** to plan a frequency **of** reporting for companies in each stratum. **NECA** tailors the reporting frequency of each stratum to reflect the significance of the data to average schedule studies. Data from strata of larger companies has a special significance because it reduces variance of sample estimates more than data fiom **strata** of smaller companies.

The five-year sampling design allows **NECA** to combine data fiom two consecutive annual samples in a single estimate without loss of effective sample size. In contrast, two consecutive samples of size 100 from each of two independent one-year sampling designs combined in an estimator would achieve a lower level of precision than two consecutive annual samples of size 100 from a single five-year sampling design with commonly defined probabilities.

In addition, **NECA** can include a larger company's data in every study while sampling it only every other year. **Thus, NECA** is able to use data that achieves the targeted precision level while sampling only half of the two-year sample each year. This feature significantly reduces costs incurred by **NECA** and by ECs, thereby reducing access charges passed on to access customers.

NECA then selects an annual sample from the five-year sample, using methods detailed in Sections II.C **through** II.G. Finally, **NECA** uses a randomization procedure to determine which study areas will be included in the sample for each of the five years. This randomization procedure assures that some companies will be selected every other year, some every third year, and some every fifth year. The reporting frequency assigned to a company is coordinated with significance of its data in average schedule studies. **This** feature assures that a greater share of the reporting costs is borne by the larger companies.

C. Sampling Design Attributes

In this section NECA describes nine attributes, which have an impact on the average schedule settlements and were used to classify the population of average schedule study areas. The 1998 Design employed nine attributes listed in Exhibit **2.1.**

With the exception of the attribute for the **size** of the company, the remaining eight attributes were used to classify the cost companies. These attributes were chosen to ensure that: (1) an adequate number of average schedule study areas were selected; **(2)** data would support development of each average schedule settlement formula with the desired level of precision; and **(3)** diverse network configurations of the universe were adequately represented.

Since there are **two** possible outcomes from each attribute, it is possible to create a total of **512** (2^9) average schedule classes. However, only 39 classes contain average schedule study areas. Similarly, the **518** cost companies populated only **52** classes out of a total possible of **256** (2^8) classes. This classification procedure created a total of 91 cost and average schedule classes. The classes created for this sampling design assure representation of the average schedule and cost company populations in terms of the relevant attributes, which have an impact on the average schedule settlements.

EXHIBIT 2.1

SAMPLE DESIGN CRITERIA

Criteria	Source/Date	
	Average Schedule	cost
1. Number of Exchanges (= 1 or > 1)	Settlement System December 1997	Customer Database December 1997
2. Size of the Company (large or small) Small: Size < 200 lines per exch.	Settlement System December 1997 Size = Access Lines/Exchanges	This criterion is not used for classifying Cost companies
3. Provider of Line Haul Facilities (yes or no)	AS 1000 Report* Line 41: Circuit Miles > 0; Line 44: Switched Circ. Terms > 0	Cost Study Database (C&WF Cat. 2 + **3** + 4 > 0) December 1997
4. Provider of Host/Remote Facilities (yes or no)	Line Haul Data Base Second Quarter 1998	Cost Study Database (C&WF Cat. **4** > 0) December 1997
5. Provider of Special Access Service (yes or no)	AS 1000 Report* Line **33:** TS Special Access Net Rev. > 0	EC1050 Report* Line **22:** Special Access Earned Rev. > 0
6. Provider of Access Tandem Facilities (yes or no)	AS 1000 Report* Line 40: ITD Settlements> 0	Cost Study Database (COE Cat. **2** > 0) December 1997
7. Traffic Volume (High or Normal) High: MPL > **325**	AS 1000 Report* $MPL = \frac{\text{Switched Access Minutes}}{\text{Access Lines}}$	Cost Study Database $MPL = \frac{\text{Switched Access Minutes}}{\text{Access Lines}}$
8. Density (High or Normal) High : Density > 175	AS 1000 Report* $Density = \frac{\text{Switched Circ. Terms.}}{\text{Exchanges}}$	EC1050 Report* $Density = \frac{\text{Switched Circ. Terms.}}{\text{Exchanges}}$
9. Participant in NECA's 1998 Traffic Sensitive Settlement Pool (yes or no)	Customer Database	**Customer Database**

A description of the 91 classes **(39** average schedule and 52 cost) with the number study areas in each of them is given in Exhibit 2.2A and 2.2B.

The columns in Exhibits 2.2A and 2.2B represent the following:

Exchanges:	Number of Exchanges
Size:	**Size** of the company
LH:	Provides Line Haul
H/R:	Provides Host/Remote
S A	F'rovides Special Access
IT:	F'rovides Tandem Switching
MPL:	Relative Access Minutes per Line
Density:	Switched Circuit Terminations per Exchange
TS:	Traffic Sensitive Pool Participant
count:	Number of Study **Areas**

EXHIBIT 2.2A

CLASSES OF AVERAGE SCHEDULE STUDY AREAS

Class	Exchanges	Size	LH	H/R	SA	IT	MPL	Density	TS	Count
1	1	large	N	N	N	N	normal	normal	N	3
2	1	large	N	N	N	N	normal	normal	Y	2
3	1	large	N	N	Y	N	normal	normal	Y	1
4	1	large	N	Y	N	N	normal	normal	N	6
5	1	large	Y	N	N	N	normal	normal	Y	6
6	1	large	Y	N	Y	N	normal	normal	Y	1
7	1	large	Y	Y	N	N	normal	normal	Y	75
8	1	large	Y	Y	N	N	high	normal	Y	4
9	1	large	Y	Y	N	Y	normal	normal	Y	1
10	1	large	Y	Y	Y	N	normal	normal	Y	138
11	1	large	Y	Y	Y	N	normal	high	Y	10
12	1	large	Y	Y	Y	N	high	normal	Y	6
13	1	large	Y	Y	Y	N	high	high	Y	3
14	1	large	Y	Y	Y	Y	normal	high	Y	3
15	1	small	N	N	N	N	normal	normal	Y	4
16	1	small	N	N	Y	N	normal	normal	Y	1
17	1	small	Y	N	N	N	normal	normal	Y	1
18	1	small	Y	Y	N	N	normal	normal	Y	12
19	1	small	*Y*	*Y*	*Y*	N	normal	nomal	Y	**3**

EXHIBIT 2.2A (Continued)

CLASSES OF AVERAGE SCHEDULE STUDY AREAS

Class	Exchanges	Size	LH	H/R	SA	IT	MPL	Density	TS	Count
20	>1	large	N	N	N	N	Normal	normal	N	2
21	>1	large	N	Y	N	N	Normal	normal	N	8
22	>1	large	Y	N	N	N	Normal	normal	Y	1
23	>1	large	Y	N	Y	N	Normal	normal	Y	8
24	>1	large	Y	N	Y	Y	Normal	normal	Y	1
25	>1	large	Y	Y	N	N	Normal	normal	Y	25
26	>1	large	Y	Y	N	N	High	normal	Y	1
27	>1	large	Y	Y	N	Y	Normal	normal	Y	7
28	>1	large	Y	Y	Y	N	Normal	normal	Y	149
29	>1	large	Y	Y	Y	N	Normal	high	Y	6
30	>1	large	Y	Y	Y	N	High	normal	Y	3
31	>1	large	Y	Y	Y	Y	Normal	normal	Y	63
32	>1	large	Y	Y	Y	Y	Normal	high	Y	14
33	>1	large	Y	Y	Y	Y	High	normal	Y	3
34	>1	small	N	Y	N	N	Normal	normal	N	1
35	>1	small	Y	Y	N	N	Normal	normal	Y	3
36	>1	small	Y	Y	N	N	High	normal	Y	1
37	>1	small	Y	Y	N	Y	Normal	normal	Y	1
38	>1	small	Y	Y	Y	N	Normal	normal	Y	2
39	>1	small	Y	Y	Y	Y	Normal	normal	Y	4
Total										583

EXHIBIT 2.2B

CLASSES OF COST COMPANY STUDY AREAS

Class	Exchanges	LH	H/R	SA	IT	MPL	Density	TS	Count
1	1	N	N	N	N	Normal	normal	N	2
2	1	N	N	N	N	Normal	normal	Y	3
3	1	N	N	Y	N	Normal	normal	Y	8
4	1	N	N	Y	N	High	normal	Y	1
5	1	Y	N	N	N	Normal	normal	N	21
6	1	Y	N	N	N	Normal	normal	Y	23
7	1	Y	N	N	N	Normal	high	N	1
8	1	Y	N	N	N	High	normal	Y	1
9	1	Y	N	Y	N	Normal	normal	Y	30
10	1	Y	N	Y	N	Normal	high	Y	2
11	1	Y	N	Y	N	High	normal	Y	4
12	1	Y	N	Y	Y	Normal	high	Y	1

EXHIBIT 2.2B (Continued)

CLASSES OF COST COMPANY STUDY AREAS

Class	Exchanges	LH	H/R	SA	IT	MPL	Density	TS	Count
13	1	Y	Y	N	N	Normal	normal	N	4
14	1	Y	Y	N	N	Normal	normal	Y	8
15	1	Y	Y	N	N	Normal	high	N	3
16	1	Y	Y	N	N	High	normal	Y	1
17	1	Y	Y	Y	N	Normal	normal	Y	11
18	1	Y	Y	Y	N	Normal	high	Y	2
19	1	Y	Y	Y	N	High	normal	Y	1
20	1	Y	Y	Y	N	High	high	Y	1
21	1	Y	Y	Y	Y	Normal	high	Y	3
22	>1	N	N	N	N	Normal	normal	N	5
23	>1	N	N	N	N	Normal	normal	Y	1
24	>1	N	N	Y	N	Normal	normal	Y	14
25	>1	N	N	Y	N	High	normal	Y	1
26	>1	Y	N	N	N	Normal	normal	N	12
27	>1	Y	N	N	N	Normal	normal	Y	5
28	>1	Y	N	N	N	Normal	high	N	1
29	>1	Y	N	N	N	High	normal	Y	2
30	>1	Y	N	N	Y	Normal	normal	N	4
31	>1	Y	N	N	Y	Normal	normal	Y	2
32	>1	Y	N	Y	N	Normal	normal	Y	38
33	>1	Y	N	Y	N	Normal	high	Y	1
34	>1	Y	N	Y	N	High	normal	Y	8
35	>1	Y	N	Y	Y	Normal	normal	Y	32
36	>1	Y	N	Y	Y	Normal	high	Y	1
37	>1	Y	N	Y	Y	High	normal	Y	4
38	>1	Y	N	Y	Y	High	high	Y	1
39	>1	Y	Y	N	N	Normal	normal	N	21
40	>1	Y	Y	N	N	Normal	normal	Y	11
41	>1	Y	Y	N	N	Normal	high	N	2
42	>1	Y	Y	N	N	High	normal	Y	1
43	>1	Y	Y	N	Y	Normal	normal	N	8
44	>1	Y	Y	N	Y	Normal	normal	Y	4
45	>1	Y	Y	N	Y	Normal	high	N	7
46	>1	Y	Y	N	Y	High	high	Y	1
47	>1	Y	Y	Y	N	Normal	normal	Y	115
48	>1	Y	Y	Y	N	Normal	high	Y	1
49	>1	Y	Y	Y	N	High	normal	Y	7
50	>1	Y	Y	Y	Y	Normal	normal	Y	65
51	>1	Y	Y	Y	Y	Normal	high	Y	9
52	>1	Y	Y	Y	Y	High	normal	Y	3
Total									518

D. Stratification of the Pouulation

NECA consolidated the 39 average schedule classes into 11 average schedule preliminary strata as shown in Exhibit 2.3A. Similarly, the 52 cost company classes were consolidated into 10 cost company preliminary strata, as shown in Exhibit 2.3B. This consolidation was based upon the number of study areas in each class and on the similarity of criteria in classes. Some of the classes listed in Exhibit 2.2A and 2.2B had too few members from which to sample and were subsequently combined with other classes. For example, classes 22 and 23 in Exhibit 2.2A were combined to form stratum A1 1 as shown in Exhibit 2.3A. Both of these classes shared common values for all attributes except traffic sensitive pool participation.

EXHIBIT 2.3A

PRELIMINARY STRATUM DEFINITION-AVERAGE SCHEDULE STUDY AREAS

Prelim. Stratum	Classes	Exch	Sue	LH	H/R	SA	IT	MPL	Density	TS	Tot.
A1	15, 16, 17, 18, 19, 34, 35, 36, 37, 38, 39	1: 12 >1: 21	small	n: 6 y: 27	n: 6 y: 27	n: 23 y: 10	n: 28 y: 5	high: 1 nrml:32	normal	n: 1 y: 32	33
A2	1, 4, 20, 21	1: 9 >1: 10	large	n	n: 5 y: 14	N	n	normal	normal	n	19
A3	8, 12, 13, 26, 30, 33	1: 13 >1: 7	large	y	y	n: 5 y: 15	n:3 y: 17	high	high: 3 nrml: 17	y	20
A4	11, 14, 29, 32	1: 13 >1: 20	large	y	Y	Y	n: 16 y: 7	normal	high	y	33
A5	9, 24, 27, 31	1: 1 >1: 71	large	y	n: 1 y: 71	n: 8 y: 64	y	normal	normal	y	72
A6	10	1	large	Y	y	Y	n	normal	normal	y	138
A7	25	>1	large	Y	y	N	..	normal	normal	y	25
A8	28	>1	large	Y	y	Y		normal	normal	y	149
A9	7	1	large	Y	y	N	n	normal	normal	y	75
A10	2, 3, 5, 6	1	large	n: 3	n	n: 8 y: 2	n	normal	normal	y	10
All	22, 23	>1	large	y	n	n:1 y: 8	n	normal	normal	y	9
Total											583

The grouping of classes causes some strata to not be completely homogeneous with regard to all of the sampling attributes. These exceptions are noted in Exhibits 2.3A and 2.3B.

For example, in stratum A1, 12 study areas have only one exchange and 21 have more than one exchange, 27 study areas have line haul facilities and 6 do not have it, 27 study areas have host remote facilities and 6 do not have it, 23 study areas do not provide Special Access services and 10 provide it, 28 study areas do not have intertoll circuits while 5 have it, one study area has high traffic volume and 32 have normal volume and all except one study area participates in the traffic sensitive pool.

Prelim. Stratum	Classes	Exch.	LH	H/R	SA	IT	MPL	Density	TS	Total
C1	1, 5, 7, 13, 15, 22, 26, 28, 30, 39, 41, 43, 45	1: 31 >1: 60	n:7 y: 84	n: 46 y: 45	n	n: 19 y:72	normal	**high:** 14 m l : 77	n	91
C2	4, 8, 11, 16, 19, 20, 25, 29, 34, 37, 38, 42, 46, 49, 52	1: 9 >1: 28	n:2 y: 35	n: 22 y: 15	n: 6 y: 31	n: 28 y: 9	**high**	high: 3 m l : 34	y	37
C3	10, 12, 18, 21, 33, 36, 48, 51	1: 8 >1: 12	y	n: 5 y: 15	y	n: 6 y: 14	normal	high	y	20
C4	31, 35, 44, 50	>1	y	n: 34 y: 69	n: 6 y: 97	y	normal	normal	y	103
C5	17	1	y	y	y	n	normal	normal	y	11
C6	40	>1	y	y	n	n	normal	normal	y	11
C7	47	>1	y	y	y	n	normal	normal	y	115
C8	14	1	y	y	n	n	normal	normal	y	8
C9	2,3,6,9	1	n: 11 y: 53	n	n: 26 y: 38	n	normal	normal	y	64
C10	23,24,27,32	>1	n: 15 y: 43	n	n: 6 y: 52	n	normal	normal	y	58
Total										518

Some preliminary strata were subdivided based on the range of interstate access revenues within the stratum. For example, the average schedule preliminary stratum A4 was subdivided into strata A4A and A4B, with total revenue <100,000 and total revenue >=100,000 respectively. Exhibits 2.4A and 2.4B show the criterion for the average schedule and cost study areas.

The average access revenue by stratum is shown in Exhibits 2.5A and 2.5B. The significant variation in the average access revenue among strata shows that this stratification effectively distinguishes companies by revenue size. For example, the average revenue for average schedule stratum A4B, is about seven times as large as that in stratum A4A.

EXHIBIT 2.4A

REVENUE SIZE CRITERION – AVERAGE SCHEDULE STUDY AREAS

Preliminary Stratum	Final Stratum	Access Revenue Criterion
A1	A1	N/A
A2	A2	N/A
A3	A3	N/A
A4	A4A	< 100,000
A4	A4B	>= 100,000
A5	A5A	< 100,000
A5	A5B	>= 100,000 & < 200,000
A5	A5C	>= 200,000
A6	A6	N/A
A7	A7	N/A
A8	A8	N/A
A9	A9	N/A
A10	A10	N/A
A11	A11	N/A

EXHIBIT 2.4B

REVENUE SIZE CRITERION – COST COMPANY STUDY AREAS

Preliminary Stratum	Final Stratum	Access Revenue Criterion
C1	C1A	< 100,000
C1	C1B	>= 100,000
C2	C2	N/A
C3	C3A	< 200,000
C3	C3B	>=200,000
C4	C4A	< 100,000
C4	C4B	>= 100,000 & < 200,000
C4	C4C	>= 200,000
C5	C5	N/A
C6	C6	N/A
C7	C7	N/A
C8	C8	N/A
C9	C9	N/A
C10	C10	N/A

EXHIBIT 2.5A

DISTRIBUTION OF ACCESS REVENUES BY FINAL STRATA

AVERAGE SCHEDULE STUDY AREAS

Stratum	Count	Average
A1	33	6,633
A2	19	69,752
A3	20	119,279
A4A	10	60,847
A4B	23	422,641
A5A	43	39,797
A5B	18	121,150
A5C	11	627,533
A6	138	27,127
A7	25	26,905
A8	149	81,629
A9	75	11,067
A10	10	16,186
A11	9	132,477

EXHIBIT 2.5B

DISTRIBUTION OF ACCESS REVENUES BY FINAL STRATA

COST COMPANY STUDY AREAS

Stratum	Count	Average Revenue
C1A	69	25,600
C1B	22	261,997
c 2	31	114,399
C3A	8	144,368
C3B	12	504,400
C4A	51	56,437
C4B	28	139,174
C4C	24	382,062
C5	11	48,092
C6	11	27,748
C7	115	83,800
C8	8	12,552
C9	64	40,656
C10	58	64,505

E. Determination of Sample Size

This section describes how NECA determined the annual sample size required to support the development of the settlement formulas. As demonstrated in previous filings, the determination is based on well-documented and widely accepted statistical sampling techniques. Sample size was determined by balancing the need to acquire reliable data against the cost and burden that such an effort places upon sampled study areas.

Experience has shown that an annual sample of approximately 100 average schedule study areas and 100 cost study areas **strikes** this balance when two consecutive annual samples are combined in each average schedule study. In order to ensure that a sufficient number of study areas are selected to account for non-response, mergers, study areas converting from average schedule to cost settlement status, and study areas exiting the NECA pools, NECA targets a higher number of study areas, about 230 per year. Of these, 115 are average schedule study areas and 115 are study areas settling on the basis of individual costs, resulting in a five-year sample size of 1150 (230 x 5).

Using data from sample companies, NECA confirmed that the resulting sample size is sufficient to provide average schedule formulas developed each year with the desired level of precision, by analyzing the precision of a sample ratio estimate of total average schedule interstate revenue requirements per access line.[7] NECA found that this ratio would be accurate within 2.5% of the true value with 95% confidence, a level sufficient for developing the average schedule formulas.

Statistical sampling textbooks, such as ***Sampling Techniques*** by William Cochran,[8] provide formulas to measure the precision of sample estimates. 'Recision" is a range about the estimate that is shown to include the true value of the universe with a designated level of confidence. NECA estimates the total average schedule revenue requirement using a stratified ratio estimate. Formulas used to calculate the precision of a stratified ratio estimate are shown below:

---

7 Total interstate revenue requirements were used in this test to ensure that the total average schedule settlements pursuant to proposed formulas would be accurate. Access line counts were **used** because **this** demand unit is the most significant determinant of total average schedule settlements. For this purpose, NECA used the April 1998 view of December 1997 data.

8 William G. Cochran, ***Sampling Techniuues***, John Wiley & **Sons**, Inc., New York, (2nd ed., 1963).

The standard error of a ratio, $\hat{R}_h$, within a stratum is given by the following formula:[9]

$$s(\hat{R}_h) = \frac{\sqrt{1-f_h}}{\sqrt{n'_h}\,\overline{X}_h}\sqrt{\frac{\sum_i (y_{i,h} - \hat{R}_h x_{i,h})^2}{N_h - 1}}$$

where:

$\hat{R}_h$ is the ratio estimate of revenue requirement per access line for stratum h.

n'_h is the size of the responding sample in stratum h. Stratum sample *sizes* are explained in Section II.F.

N_h is the number of study areas in stratum h.

$x_{i,h}$ is the number of access lines for study area i in stratum h, and is taken from the April 1998 view of December 1997 data.

$y_{i,h}$ is the total interstate revenue requirement for study area i in stratum h, and is taken from the April 1998 view of December 1997 data.

f_h is the ratio of the responding two-year sample size in stratum h (n'_h) to the total number of study areas (N_h) in stratum h.

$\overline{X}_h$ is the mean of access lines for **stratum** h displayed in Column H of Exhibit 2.7.

In this formula, the value $\hat{R}_h$ and the summation are calculated using data from all study areas in each stratum h.

Exhibit ***2.6*** shows an example of the calculation of the standard error and variance of the ratio estimate for average schedule stratum A2. Study areas in **this** exhibit correspond to those in average schedule stratum ***A2*** in Appendix A1. Columns B, C and D show the calculation of components of $\hat{R}_{A2}$. Column E shows the calculation of the sum of squares component of the variance.

9 *Id.* at **p.31.**

EXHIBIT 2.6

REVENUE REOUIREMENTS AND ACCESS LINES FOR AVERAGE SCHEDULE STRATUM *A2*

(A) Study Area Observation No.	(B) Revenue Requirement	(C) Access Lines	(D) $\frac{\Sigma(Col.B)}{\Sigma(Col.C)}$	(E) $((B)-(D)(C))^2$
	(y_{ih})	$(x_{i,h})$	$(\hat{R}_h)$	$(y_{i,h} - \hat{R}_h x_{i,h})^2$
1	15,283	2.03 1		140,898.17
2	20.543	2.765		61,389 66
^	20,977	2,724		965,721.53
4	27.025	3.568		698,413.61
	27,315	3,710		6,959.69
6	32.224	4.395		1.260.79
7	34,052	4,447		1,990,384.62
8	38.616	5.287		36,416.00
o	39,309	5,079		4,116,435.45
10	48.097	6,610		176.998.81
11	54,747	7,538		339,045.47
12	65.506	8.830		4.576.913.25
13	71,286	8,633		41,904,685.58
14	86.936	11.725		763,781.77
15	107,684	14,925		3,482,721.42
16	145.518	20.103		4.157.397.45
17	256,718	35,417		10,526,476.33
18	259.371	36.092		30,757,950.19
19	370,313	50,659		2,330,059.44
TOTAL	1,721,520	234,538	7.34	107,033,909.22

$$\hat{R}_{A2} = \frac{1,721,520}{234,538} = 7.34$$

$$s(\hat{R}_{A2}) = \frac{\sqrt{1-0.52632}}{(\sqrt{10})\,(12,344.11)}\sqrt{\frac{107,033,909.22}{19-1}} = 0.042426$$

$$Var(\hat{R}_{A2}) = s(\hat{R}_h)^2 = (0.042426)^2 = 0.0019$$

Exhibit 2.7 shows the resulting variance of the ratio estimate for each stratum. Column C shows the resulting stratum variances. The stratum variances were then used to determine the variance of the overall stratified ratio estimator, $\hat{R}$, using the following formula:[10]

$$Var(R) = \frac{\sum_{h=1}^{L} X_h^2 Var(R_h)}{X^2}$$

Where: X_h is the total of access lines in stratum **h.**

X is the total of population access lines.

Columns B, C and D of Exhibit 2.7 show the components of this calculation.

[10] *Id.* at **p.** 90. Formula 5.3 found in <u>*Sampling Techniques*</u> note ***6 supra*** is ***a*** similar expression. NECA used the sum of access lines as the weighting factor.

EXHIBIT 2.7

AVERAGE SCHEDULE STRATUM VARIANCE DATA

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)=(B)/(E)
Stratum	**Access Lines**	$\mathrm{Var}(\hat{R}_h)$	**$(B)^2$ x (C)**	N_h	n'_h	**(F)/(E)**	**Mean Access Lines**
A1	11,406	8.9328	1,162,128,528	33	13	0.39	345
A2	234,538	0.0019	101,682,214	19	10	0.53	12,344
A3	71,146	0.0000	0	20	20	1.00	3,557
A4A	32,476	7.1118	7,500,706,165	10	4	0.40	3,247
A4B	484,159	0.0000	0	23	23	1.00	21,050
A5A	105,359	6.6424	73,734,507,772	43	12	0.28	2,450
A5B	120,935	3.5861	52,447,111,668	18	4	0.22	6,718
A5C	433,309	0.0000	0	11	11	1.00	39,391
A6	223,769	0.8352	41,819,379,046	138	34	0.25	1,621
A7	43,153	12.9344	24,086,151,806	25	5	0.20	1,726
A8	689,543	0.3693	175,571,616,270	149	74	0.50	4,627,81
A9	55,540	9.7597	30,105,541,420	75	9	0.12	740
A10	10,333	1.0843	115,768,272	10	4	0.40	1,033
A11	72,496	4.7161	24,786,280,412	9	4	0.44	8,055
Total	2,588,162		431,430,873,573				

Using values from the exhibit, the overall variance of the ratio estimate is calculated as follows:

$$\mathrm{Var}(\hat{R}) = \frac{431{,}430{,}873{,}573}{(2{,}588{,}162)^2} = 0.064406$$

NECA then developed a 95% confidence interval to determine the relative precision of the estimator, using the formula below.”

$$\Pr\left(|\hat{R} - R| \geq d\right) = 0.05$$

Or

$$d = \frac{Z_{.05} \times s(\hat{R})}{R}$$

where:

$Z_{.05}$ is the value of standard **normal** distribution N(0,1) corresponding to 95% confidence level. which **is** 1.96.

d is the difference between the estimated and true value of R.

R is the **ratio** of revenue requirements to **access** lines for the entire population of average schedule study areas for December 1997, a value of 19.743137

Substituting data results in the following:

$$d = \frac{1.96 \times \sqrt{0.064406}}{19.743131} = 0.0252\%$$

This calculation shows that the average schedule sample is precise within 2.52% at the 95% confidence level, a level sufficient for average schedule development.

F. Allocation of Sample to Strata

NECA allocated the total sample size to strata using a method, known as “Neyman Allocation”, a method which produces optimum precision results for stratified sampling.” The Neyman Allocation

11 *Id.* at p. 75.

12 *Id.* at **p.** 97.

determines the size of the sample in each stratum in proportion to an estimate of the standard deviation of a measure of size in each stratum. The Neyman allocation is optimum (improves precision most) when the measure of size is correlated with the variable to be estimated (revenue requirement). The Neyman allocation to a stratum also depends upon the total count of study areas in the stratum (Column C of Exhibit 2.8A and 2.8B), and the number of study areas in the five-year sample. Following are derivations of these standard deviations and the count of study areas in the sample.

NECA defined a study area's measure of size to be the square root of its total interstate access revenues for two reasons. This measure relates to the variation in revenue requirements among average schedule companies, and it reduces the likelihood of over-allocation to strata of large study areas that would result from use of a measure of size that did not use the square root. These values are shown in Appendix A1.

Next, the standard deviation of measure of size in each stratum is calculated. These values are shown in Column B of Exhibits 2.8A and 2.8B. For example, for average schedule stratum A1, the standard deviation of the measure of size is 49.85.

The total five year sample size of 1150 was allocated in the following steps.

1. Study areas in strata with high traffic volume (> 325 minutes per line, strata A3, C2) were designated to be censused and sampled every other year.
2. The remaining sample size were allocated using the Neyman Allocation.
3. Each allocation was tested to assure that no study area would be sampled more often than every other year. Strata with sample size allocations larger than this were also censused and sampled every other year.

4. The remaining sample size was allocated according to the Neyman Allocation.

By this method, strata A3, A4B, A5C, C2 and C3B were censused and the remaining total trial five-year sample size of 893 $(1150 - \frac{103x5}{2})$ was allocated according to the Neyman Allocation.

Exhibits 2.8A and 2.8B show the use of standard deviations and the total trial five year sample size to calculate trial stratum five year sample sizes for average schedule and cost companies.

The sample allocation weight (column D) is calculated as the product of the standard deviation of the measure of size (column B) and the number of study **areas** (column C). The sample allocation weight for a particular stratum, divided by the **sum** of all sample allocation weights, produces a stratum allocation fraction. This fraction was multiplied by the total trial five-year sample size to produce a trial five-year sample size in each stratum (Column E).

$$\text{sample allocation weight}_h = SD_h(MOS) \times N_h$$

$$\text{sample allocation fraction}_h = \text{allocation weight}_h / \Sigma(\text{allocation weight}_h)$$

$$\text{trial stratum five-year sample size} = \text{sample allocation fraction}_h \times \text{total trial five-year sample size}$$

For example, for average schedule stratum ***A2*** in Exhibit 2.8A, the trial stratum five-year sample size is calculated **as** follows:

$$\text{sample allocation weight}_{A2} = 123.347 \times 19 = 2343.59$$

$$\text{sample allocation fraction}_{A2} = 2343.59 / 87704.42 = 0.0267215$$

$$\text{trial stratum five-year sample size}_{A2} = 0.0267215 \times 892.5 = 23.849 \quad (-24)$$

The trial stratum annual sample size is calculated as the integer part **of** $\frac{\text{trial stratum five year sample size}}{5} + 0.5$.

The sampling term, which represents how often a study area will be sampled, is calculated as the integer part of $\frac{N_h}{\text{tnal annual sample size}} + 0.5$, but is limited to a value between 2 and 5.

The final five-year sample size is calculated to be the integer part of $(\frac{N_h}{\text{Sampling Term}} \times 5)$. It is adjusted to be at least 10.

For example, continuing the calculation for average schedule stratum ***A2*** in Exhibit 2.8A,

$$\textbf{\textit{trial annual Sample Size}} = \text{int}(\frac{23.849}{5} + 0.5) = 5$$

$$\textbf{\textit{sampling term}} = \text{int}(\frac{19}{5} + 0.5) = 4$$

$$\textbf{\textit{final five-year sample size}} = \text{int}(\frac{19}{4} \times 5) = \text{int}(23.75) = 23$$

$$\textbf{\textit{final annual sample size}} = [5,4] \text{ (i.e., alternating 5 in the first, 4 in the second year)}$$

EXHIBIT 2.8A

FINAL STRATA – AVERAGE SCHEDULE STUDY AREAS

(A) Stratum No.	(B) Standard Deviation of M. O. S.[13]	(C) No. of Study Areas	(D) Sample Allocation Weight	(E) Trial Five Year Sample Size	(F) Sampling Term	(G) Final Annual Sample Size	(H) Final Five Year Sample Sue
A1	49.85	33	1645.08	17	5	[7,6]	33
A2	123.35	19	2343.59	24	4	[5,4]	23
A3	199.61	20	0.00	0	2	10	50
A4A	55.30	10	552.95	6	5	2	10
A4B	308.23	23	0.00	0	2	[12,11]	58
A5A	71.00	43	3053.00	31	5	[6,5]	30
A5B	25.33	18	455.89	5	5	2	10
A5C	444.04	11	0.00	0	2	[6,5]	28
A6	65.04	138	8976.07	91	5	17	86
A7	55.89	25	1397.30	14	5	[3,2]	15
A8	122.47	149	18248.03	186	4	[38,37]	186
A9	31.54	75	2365.35	24	5	[5,4]	25
A10	34.22	10	342.17	3	5	2	10
A11	106.71	9	960.38	10	5	[2,1]	10
TOTAL		583	40339.82	411		[108] to [116	574

13 M. O. S. is "Measure of Size.'' It is the square root of access revenues.

EXHIBIT 2.8B

(A) Stratum No.	(B) Standard Deviation of M. O. S.	(C) No. of Study Areas	(D) Sample Allocation Weight	(E) Trial Five Year Sample Size	(F) Sampling Term	(G) Final Annual Sample Size	(H) Final Five Year Sample Size
C1A	78.68	69	5,429.13	55	5	[11,12]	57
C1B	148.04	22	3,256.88	33	3	[7,8]	36
C1C	128.49	58	7,452.48	76	4	[14,15]	72
C2	179.87	37	0.00	0	2	[18,19]	93
C3A	71.42	8	571.34	6	5	[1,2]	10
C3B	218.38	12	0.00	0	2	6	30
C4A	55.56	51	2,833.46	29	5	[5,6]	28
C4B	38.05	28	1,065.37	11	5	2	10
C4C	142.91	24	3,429.77	35	3	8	40
C5	78.74	11	866.11	9	5	[1,2]	10
C6	47.02	11	517.21	5	5	[1,2]	10
C7	110.22	115	12,675.30	129	4	[28,29]	143
C8	48.17	8	385.38	4	5	[1,2]	10
C9	138.79	64	8,882.24	90	4	16	80
TOTAL		518	47364.67	482		[119] to 129]	629
GRAND TOTAL[14]		1101	87704.42	893			1203

[14] The Grand Total is the sum **of** the Totals from Exhibits 2.8A and 2.8B. The Sample Allocation Weight Grand Total is used to calculate Column E.

G. Selection of Samule

In this section, NECA describes methods for selecting sample study areas. To obtain reliable estimates from a sample requires that each member of the population has a well-defined probability of inclusion in the sample. NECA chose a particular method of defining probabilities because it produces greater precision than other methods.

NECA determined the probability of including a specific study area in the five-year sample using one of two methods. Study areas in the census strata" were assigned a probability of one for inclusion in the multi-year sample. Study areas from other sample strata were assigned probabilities proportionate to size (PPS). The PPS method was used because it provides more precise estimates than do other probability sampling methods.

Calculations supporting the PPS method are detailed in Appendix A1. Study areas within a stratum are ordered, according to their measure of size, starting with the largest. For example, in cost stratum C4B study area number one has the highest measure of size (443.5). Next, the cumulative measure of size is computed as a running total of measures of size. The cumulative measure of size associates a range of measure of size values with each study area, including all values between the study area's cumulative measure of size and the cumulative measure of size of the preceding study area. For example, the range of measure of size associated with study area one in cost stratum C4B is 0 to 443.5. Similarly, the range of size associated with the next study area is from 443.5 to 877.14.

The stratified PPS method divides each stratum into sampling intervals, then selects one sample member from each interval. The sampling interval is determined by dividing the stratum total

---

[15] Specifically, cost strata C2, C3B and average schedule strata A3, A4B, A5C

measure of size by the stratum five-year sample size reported in column H of Exhibit 2.8. For example, in cost stratum C4B, the stratum sampling interval is:

$$\textbf{\textit{Stratum Sampling Interval}} = \frac{10393.2}{10}$$

$$= 1039.32$$

The PPS method selects sample members from intervals systematically, selecting the first member by a random start, then successively adding an interval to the random start to select other sample members. The random start for each stratum was computed by multiplying a random number by the stratum sampling interval. Random starts calculated by this method are displayed in Exhibit 2.9.[16]

In each stratum, the sample study area whose Measure of Size range included the stratum's random start was selected. A sequence of sample selection numbers was identified by progressively adding the stratum sampling interval to the random start. Each study area whose measure of size range included one of these values was included in the multi-year sample. For example, for cost stratum C4B shown in Appendix A1, this method first selects the study area with sequence number 3 because the random start for this stratum (955.46) is within study area 1 range of measure of size, which extends from 877.14 to 1308.69. Similarly, study area 8 is included in the sample because by calculating a second random number in the stratum (random start+ $2\times$ sampling interval = 955.46 + 2078.64 = 3034.1), it is determined that 3034.1 is within the study area 8 range of measure of size.

16 Random numbers were generated using the RANUNI function of the SAS computer software. The function returns a number generated from the uniform distribution on the interval [0,1] using a prime modulus multiplicative generator with modulus 2^{31} - 1, and multiplier 397,204,094. ***See*** **SAS** Institute, *SAS Language:* ***Reference, Version*** 6,592 (1st ed. 1990).

Results for all strata are displayed in Appendix A1.

EXHIBIT 2.9

RANDOM STARTS FOR EACH STRATUM

Stratum	Random Start
A1	0.506911
A2	0.558424
A3	0.444686
A4A	0.844831
A4B	0.277616
A5A	26.194459
A5B	250.336799
A5C	0.3911658
A6	71.611490
A7	176.341963
A8	0.841177
A9	299.349308
A10	0.015482
A11	0.017919
C1A	44.343804
C1B	0.883257
C1C	0.761835
C2	0.009090
C3A	0.907439
C3B	0.597745
C4A	407.681308
C4B	955.464763
C4C	0.882367
C5	127.413264
C6	152.189704
C7	0.897793
C8	0.460138
C9	0.232646

When a sample is selected by this method, the probability that a particular study area is included in the five-year sample is:

$$\textit{Probability of Inclusion in the Five-Year Sample} = \frac{\textit{Stratum Five-Year Sample Size} \times \textit{Study Area Measure of Size}}{\textit{Total Stratum Measure of Sue}}$$

For example, for study area one within cost stratum C4B,

$$\textit{Probability of Inclusion in the Five-Year Sample} = \frac{10 \times 443.5}{10393.2} = 0.42672$$

According to this formula, large study areas have a higher probability of inclusion than do smaller ones. In cases where this formula would produce a value greater than one, a probability of inclusion equal to one was assigned.

The Probability of Selection ... a particular year sample is given by:

$$\textit{Probability of Selection} = \frac{\textit{Probability of Inclusion in Five-Year Sample}}{\textit{Stratum Sampling Term}}$$

For example, the probability of selecting Study Area 1 within Stratum C4B in any given year is:

$$\textit{Probability of Selection} = \frac{0.42672}{5} = 0.085344$$

H. Samule Weights

In all probability samples, each member of the sample represents a determined share of the population. For example, in a simple random sample of 5 out of 50, each sample member represents

10 population members, and so has a probability of selection equal to 0.1. To derive an estimate of the population total from such a sample, we would multiply the sample total by 10. In **this** case, 10 would be the sample weight, applied equally to each member of the simple random sample. In a probability sample which is not a simple random sample, probabilities of selection are unequal. Correspondingly, sample weights are unequal and are unique for each member of the sample. Each sample weight is the reciprocal of the probability of selection:

$$\textbf{\textit{Sample Weight}} = \frac{I}{\textbf{\textit{Probability of Selection}}}$$

For example, when using data from study area 1 **within** cost stratum C1A, as part of a single year sample to estimate a population total, the sample weight would be:

$$\textbf{\textit{Sample Weight}} = \frac{1}{0.085344} = \mathbf{11.72}$$

NECA's studies combine data from two consecutive samples. Consequently, probabilities of inclusion in the double sample are twice the probability of selection in the one-year sample. Therefore, the sample weights **used** by NECA with the double sample equal one-half the one-year sample weights.

I. Assignment of Study Areas to Samule Years

This section describes how study areas selected for inclusion in the five-year sample are assigned to at least one, and to **as** many **as** three years of the **five** sample years.

Column F of Exhibits **2.8A** and **2.8B** specify the sampling term assigned to each stratum. **A** sampling term of three, for example, means that a company selected in the 1999 sample would be selected next for the **2002** Sample, or every third year. Shorter sampling terms were assigned to strata consisting of larger study areas, while longer sampling terms were assigned to strata consisting of smaller study areas. For example, in Exhibit **2.8A,** cost stratum C1A was assigned a sampling term of five, while cost stratum C1B was assigned **a** sampling term of three.

To make this assignment, for each stratum, a list of consecutive integers was assembled in random order, which counts from 1 to t, where t **is** the stratum sampling term. For example, in cost stratum C1**A** (which has a sampling term of five), the first random number was 1, followed in sequence by **4, 2, 3,** and **5.** Next, these randomly ordered numbers were assigned consecutively to sample study areas. Study areas, which were assigned a random number equal to 1 are sampled in the first year; those with a number equal to **2** are sampled in the second year, etc.

In strata with sampling terms less than 5, study areas are repeated in random number order in sample years after the term is reached. For example, in a stratum with a term of **2,** a study area with a random number equal to 1 would also be sampled in the third and fifth year.

The annual sample size for each stratum, which was produced by this randomization method, is shown in Column G of Exhibits **2.8A** and Exhibit **2.8B**. In some strata, the sample sizes are not the same in every year because the multi-year sample size did not divide evenly by the term. In such cases, numbers in parenthesis designate the alternating year sample sizes.

Thus, the current five-year sample design accurately and efficiently represents the total average schedule population. Methods described herein assure that sample data represent the costs of each settlement function, for large and small companies, having normal, low and high cost conditions.